September 2002: Use Text For a Few Numbers, Tables for Many Numbers, Graphs for Complex Relationships (Rule 7.1)

Rules of the month are numbered in accordance with the numbering in the book. Thus, Rule 1.1 refers to the first rule in Chapter 1. And so on. These comments do not repeat the material in the book but highlight and amplify it. A rule is stated—as found in the book—and then discussed.

Statement of Rule 7.1

"Use sentence structure for displaying 2 to 5 numbers, tables for displaying more numerical information, and graphs for complex relationships."

Tables vs. Graphs

The question we face is this discussion is: "Are tables appropriate for showing complex relationships?" The answer of Rule 7.1 is obviously, "No!" But when does a relationship become "complex"? (Speaking statistically, of course.) In a very well-done paper Gelman, Pasarica, and Dodhia (2002) [abbreviated as GPD] argue that statisticians overuse tables to summarize and associate. They went systematically through one issue of the *Journal of the American Statistical Association* (March 2000), analyzed all the tables presented, and gave graphical alternatives for a subset of tables that, they argued, were more illustrative and trenchant. I think they were right.

Some pros and cons for the use of tables.

1. Tables are more appropriate when there are many "localized comparisons," to use Tufte's (1983) terminology. Thus, the first table in an epidemiological article describing the sample, cohort, or population deals with many such comparisons and it is difficult to put it into a graph. Similarly, the base frequencies of a questionnaire usually are best put into a table. (Of course, relating items on the questionnaire may be done more appropriately by means of graphs). Also, as I mentioned in the text, census data is better tabulated rather than graphed.
2. There is a selfish reason for preferring tables: the data can be used for teaching purposes. There are many times that I wished I had had the raw data of a graph in order to do my own analysis. Occasionally I have been so desperate that I've read the values off the graph!
3. Numerical simulations (a staple of statistical journals) are grist for the graphics mill since the actual values are usually of lesser interest than the patterns and interrelationships. Several of the examples by GPD deal with these kinds of situations. Such tables are characterized by a degree of repetitiveness. For example, the Type I error associated with four methods of estimation and three different sample sizes.

4. A graph is more challenging to construct than a table. In part because it forces the writer to be more purpose driven. A picture may be worth one thousand words. but the words usually tell a story. And the nature of a story is that it is unique. If you tell Story A, then you cannot tell Story B. Hence graphs are great for purpose-driven articles. (That is another reason that census data is usually tabulated rather than graphed: it's the user that finds the story, rarely the Census Bureau.)
5. Data mining is a form of statistical story-telling ranging from outright fiction to utter reality. Data miners use lots of graphs and pictures to illustrate their finds.
6. A graph does not need to take much more space than a table. As GPD point out, a graph may take more explanation.
7. Table 7.5 in *Statistical Rules of Thumb* is a good illustration of all the above. The table is graphed in two ways in Figure 7.4. The first as a bar graph, the second as a more meaningful line graph. Below are the data of Table 7.5. Before looking at the book, try to graph these data in some meaningful way. Then go to Figure 7.4 and compare. If you have another way of presenting these data let me know and I will add to the response section.

Number of social activities in a two-week period among persons 70 years of age and older. Table 7.5 from *Statistical Rules of Thumb*; data cited are from Kramarov et al. (1999).

	Age			
Number of Activities	70-74 years	75-79 years	80-84 years	85 years and over
	%	%	%	%
Women				
0 activities	1.0	1.3	2.1	3.1
1-2 activities	6.8	10.5	11.9	19.2
3-4 activities	26.8	27.5	32.5	38.3
5-7 activities	65.4	60.7	53.5	39.4
*Mean number**	*4.96*	*4.76*	*4.53*	*3.99*
Men				
0 activities	1.9	1.7	2.9	5.3
1-2 activities	10.5	13.3	15.9	23.0
3-4 activities	26.3	30.3	36.7	35.9
5-7 activities	61.2	54.7	44.5	35.9
*Mean number**	*4.75*	*4.54*	*4.17*	*3.96*

*Data calculated from percentages; not in the original table.

References

Gelman, A., Pasarica, C. and Dodhia, R. (2002). Let's practice what we preach: turning tables into graphs. *The American Statistician,* **56:** 121-130.
Tufte, E.R. (1983). *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, CT.
Kramarov, E., Lentzner, H., Rooks, R., Weeks, J. and Saydah, S. (1999). *Health and Aging Chartbook. Health, United States, 1999.* National Center for Health Statistics, Hyattsville, MD.
(http://www.cdc.gov/nchs/data/hus/hus99.pdf).

Responses

This section is intended to contain reader comments and perhaps responses from me. It provides a forum for discussion and further reflection.